STA522 Project2

Holly Cui, Elyse McFalls

2023-12-04

Overview

While paper helicopters may seem like child's play, have you ever pondered the strategic nuances behind crafting the optimal design to conquer the skies? In the realm of whimsical flights and childhood contests, we dissect the intricacies of paper helicopter design, armed with the fundamental tools of inquiry – regular paper, a pair of scissors, and a humble paperclip, to explore the relationship between paper helicopter features and flight duration. Specifically, we focuses on four major factors in helicopter designs that are commonly assumed to effective – the rotor length, the leg length, the leg width, and the paperclip-on-leg maneuver. By altering the combinations of these factors, we aim to answer the following research questions:

1. Which factors seem to be the most important for making helicopters that fly longer (in terms of time)?
2. Is there any evidence that the effect of rotor length differs by leg width?
3. What would be recommended as the ideal combination to make the helicopter fly long (in terms of time)?

Methodology

In this study, we performed a full factorial randomized experiment on all 2^4 combinations of the main factors. Each combination is sampled 5 flights, leading to a total of 80 observations as our sample size. During the sampling process, the treatment values are randomly permuted under the guideline of a randomized experiment, and the flight duration is sequentially collected by dropping the assigned combination of paper helicopter from a fixed height of 6'6''.

Upon designing the experiment, we believe that the full factorial methodology presents the best balance between reliability and convenience. Since our research questions are interested in determining the best combination of all four factors, we wanted to make sure that the main effects of each factor as well as the interaction effects can be estimated without further assumptions. Thus, the alternative of a fractional factorial design would either not allow us to calculate all effects due to aliasing, or require tedious planning on how to optimally craft the design. Considering 2^4 combinations is an acceptable level of time commitment on data collection, we decided to pursue on the full factorial design.

For reference, the 16 conditions of paper helicopters are defined as follows:

ID	Treatment	ID	Treatment
a	Only high rotor length	bd	High leg length & leg clip
b	Only high leg length	cd	High leg width & leg clip
c	Only high leg width	abc	High rotor length & high leg length & high leg width
d	Only leg clip	abd	High rotor length & high leg length & leg clip
ab	High rotor length & high leg length	acd	High rotor length & high leg width & leg clip

ID	Treatment	ID	Treatment
ac	High rotor length & high leg width	bcd	High leg length & high leg width & leg clip
ad	High rotor length & leg clip	abcd	High on all factors
bc	High leg length & high leg width	1	Low on all factors

where the levels of highs and lows are defined as:

- Rotor length: low = 7.5 cm, high = 8.5 cm
- Leg length: low = 7.5 cm, high = 12.0 cm
- Leg width: low = 3.2 cm, high = 5.0 cm
- Leg clip: no, yes

The randomizations of treatments are generated in R:

```
set.seed(123)
trt = c(rep("a", 5), rep("b", 5), rep("c", 5), rep("d", 5),
        rep("ab", 5), rep("ac", 5), rep("ad", 5), rep("bc", 5),
        rep("bd", 5), rep("cd", 5), rep("abc", 5), rep("abd", 5),
        rep("acd", 5), rep("bcd", 5), rep("abcd", 5), rep("1", 5))
sample(trt)
```

```
##  [1] "ad"   "1"    "abc"  "c"    "bcd"  "bd"   "cd"   "bd"   "1"    "ab"
## [11] "bcd"  "abd"  "b"    "ac"   "b"    "abcd" "bcd"  "acd"  "d"    "bc"
## [21] "abcd" "d"    "abcd" "bc"   "abc"  "c"    "c"    "ad"   "acd"  "bd"
## [31] "bcd"  "acd"  "bd"   "b"    "ab"   "ac"   "abcd" "1"    "abd"  "bc"
## [41] "abcd" "ad"   "ac"   "a"    "b"    "abc"  "c"    "d"    "a"    "bc"
## [51] "abd"  "1"    "b"    "ab"   "abc"  "cd"   "1"    "d"    "d"    "ac"
## [61] "c"    "acd"  "bd"   "cd"   "abc"  "ac"   "a"    "abd"  "abd"  "a"
## [71] "bc"   "ab"   "ad"   "a"    "bcd"  "ab"   "ad"   "cd"   "cd"   "acd"
```

Here is a sneak peek of our final dataframe, where variable `time` is measured in seconds, variable `trt` represents the treatment combinations, and variables `rotor`, `leg_len`, `leg_wid`, and `clip` are binary indicators of the presence of high/yes (=1) or low/no (=0):

```
plane = read.csv("paperplane.csv", header = T)
kable(head(plane), booktabs = T, format = "latex") # dimension: 80 rows 6 columns
```

time	rotor	leg_len	leg_wid	clip	trt
2.03	1	0	0	1	ad
1.99	0	0	0	0	1
1.70	1	1	1	0	abc
1.93	0	0	1	0	c
1.32	0	1	1	1	bcd
1.68	0	1	0	1	bd

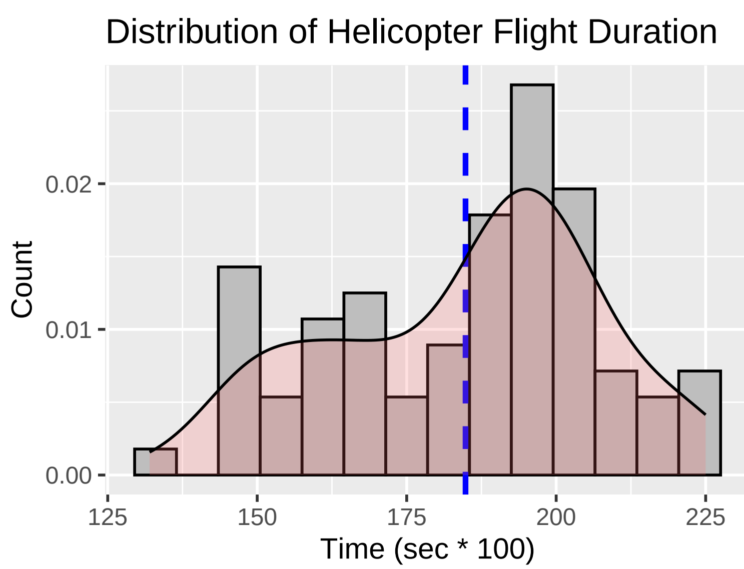
Exploratory Data Analysis

```
plane <- plane %>%
  mutate(rotor = as.factor(rotor),
         leg_len = as.factor(leg_len),
         leg_wid = as.factor(leg_wid),
         clip = as.factor(clip))
```

Outcome Distribution

As a functional check, we visualized the distribution of our collected flight duration in seconds magnified by 100 times to gain some insights. The plot below exhibits a roughly bimodal distribution with peaks around 1.6 seconds and 1.9 seconds. The overall shape is also slightly right-skewed, with an average time of 1.85. It is yet to be decided whether the normality concern should be fixed based on the marginal distribution of Y|X (visualized below as a boxplot) along with the model diagnostic plots on residuals, but we will not worry too much as of the small sample size in each treatment.

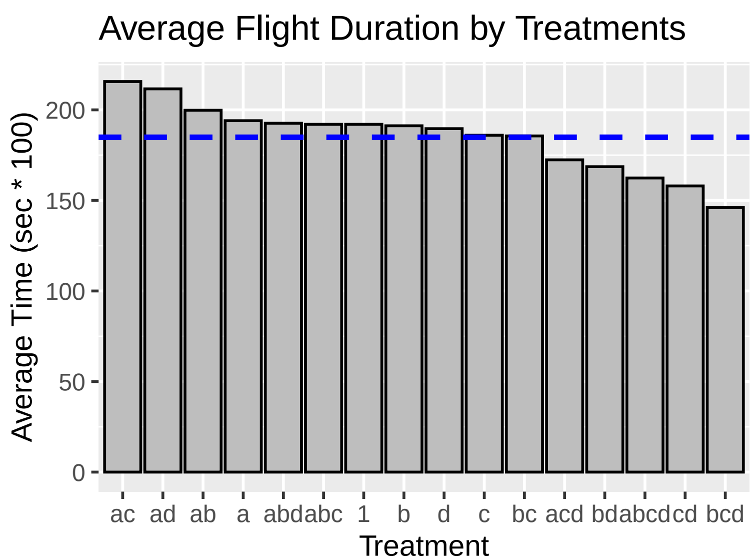
```
ggplot(plane, aes(x = time*100)) +
  geom_histogram(aes(y=after_stat(density)),
                 binwidth=7, color="black", fill="grey") +
  geom_vline(aes(xintercept = mean(time)*100),
             color="blue", linetype="dashed", linewidth=1) +
  geom_density(alpha=.2, fill="#FF6666") +
  labs(title = "Distribution of Helicopter Flight Duration",
       x = "Time (sec * 100)", y = "Count")
```



Outcome by Treatments

We proceed to explore the general behaviors of flight duration by treatments. Looking at the average time per treatment combination, we see that helicopters with high rotor length and high leg width (ac) last the longest in the air. Moreover, the top six combinations all contain high rotor length (a), indicating that it may contribute to longer flight times. Conversely, helicopters with high leg length, high leg width, and a leg clip (bcd) are the quickest to go down. Interestingly, this combination is high on all levels except rotor length. And six of the eight worst combinations by average flight time have low rotor length and leg clips (d).

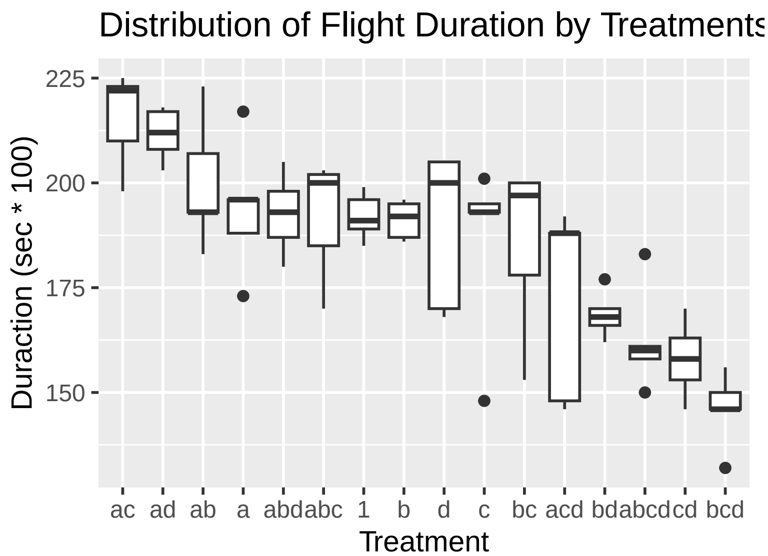
```
plane %>%
  group_by(trt) %>%
  summarise(mean_time = mean(time*100)) %>%
  arrange(desc(mean_time)) %>%
  ggplot(aes(y = mean_time, x = fct_rev(fct_reorder(trt, mean_time)))) +
  geom_col(color="black", fill="grey") +
  geom_hline(aes(yintercept = mean(mean_time)),
             color="blue", linetype="dashed", linewidth=1) +
  labs(title = "Average Flight Duration by Treatments",
       x = "Treatment", y = "Average Time (sec * 100)")
```



If we proceed to the boxplot showing the flight duration distribution by each treatment group, it is relatively difficult to interpret with only 5 observations per box. However, some insights can be observed:

- There are several outliers in treatment bcd and other groups. We may want to estimate the model with and without that point to see if our conclusions are overly sensitive to that point. But since most outliers in groups are within the normal range, plus the actual difference is measured on millisecond scales, we will not worry too much.
- The distribution of time in each box seems normal, except for some groups such as ac, a, and d, where the right-skewed shape persists. Thus, we might expect some points in our modeling exhibiting non-normal behavior with respect to residuals, but we would not carry out further transformations for the sake of interpretability.
- The variance for group d and acd are relatively large. However, since the total duration is measured under second, the largest difference is only milliseconds away, so we could still proceed with regression under the equal variance assumption. Moreover, from the actual variance calculations, most of the groups confirmed to be the same. With such small sample size, it is difficult to conclude definitively that the variances are radically different across groups.

```r
plane %>%
  mutate(trt = factor(trt, levels = c("ac", "ad", "ab", "a",
                                      "abd", "abc", "1", "b",
                                      "d", "c", "bc", "acd",
                                      "bd", "abcd", "cd", "bcd"))) %>%
ggplot(aes(x = trt, y = time*100)) +
  geom_boxplot() +
  labs(title = "Distribution of Flight Duration by Treatments",
       x = "Treatment", y = "Duraction (sec * 100)")
```




```r
tapply(plane[,1], plane[,6], sd)
```

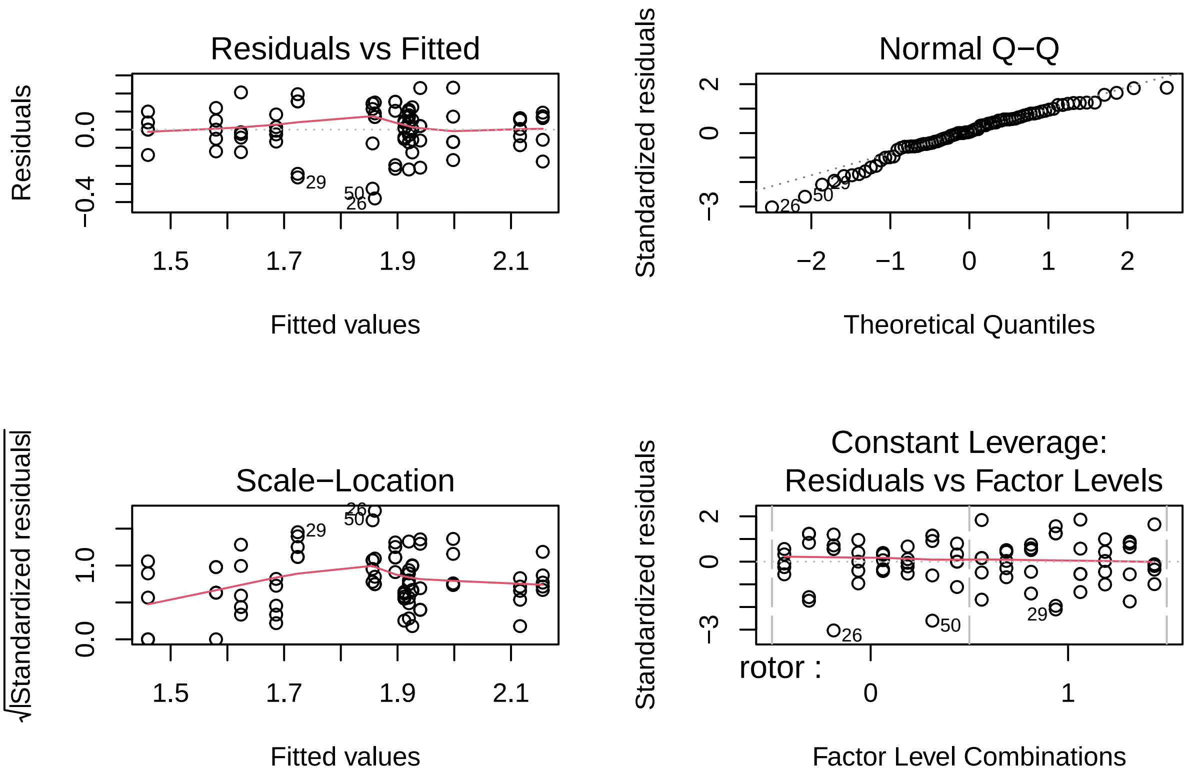
```
##          1          a         ab        abc       abcd        abd         ac 
## 0.05567764 0.15921683 0.15530615 0.14300350 0.12300406 0.09659193 0.11458621 
##        acd         ad          b         bc        bcd         bd          c 
## 0.23255107 0.06268971 0.04549725 0.20403431 0.08831761 0.05549775 0.21494185 
##         cd          d 
## 0.09192388 0.18928814
```

Modeling

Q1: Most Important Factor for Longer Flight

To assess the relative importance of each factor, we first adopt a linear regression model under the assumption that all interactions between factors exist. Checking model diagnostic plots, we can say that the model assumptions are not violated and the residual distribution is aligned. However, simply comparing the coefficient estimates from model output would not render us a statistically reliable answer.

```r
mod.all = lm(time ~ rotor * leg_len * leg_wid * clip, data = plane)
par(mfrow = c(2,2))
plot(mod.all)
```



In order to compare the contribution of each factor, we decided to use multiple nested F-tests. In each comparison, two models with and without a specific factor are subject to testing. After collecting all test statistics, we chose the best factor yielding the most significant results.

Rotor Length

```
mod.without_a = lm(time ~ leg_len * leg_wid * clip, data = plane)
anova(mod.without_a, mod.all)
```

```
## Analysis of Variance Table
##
## Model 1: time ~ leg_len * leg_wid * clip
## Model 2: time ~ rotor * leg_len * leg_wid * clip
##   Res.Df    RSS Df Sum of Sq     F    Pr(>F)
## 1     72 1.8876
## 2     64 1.2547  8   0.63285 4.035 0.0006263 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Leg Length

```
mod.without_b = lm(time ~ rotor * leg_wid * clip, data = plane)
anova(mod.without_b, mod.all)
```

```
## Analysis of Variance Table
##
## Model 1: time ~ rotor * leg_wid * clip
## Model 2: time ~ rotor * leg_len * leg_wid * clip
##   Res.Df    RSS Df Sum of Sq    F  Pr(>F)
## 1     72 1.6641
## 2     64 1.2547  8   0.40935 2.61 0.01547 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Leg Width

```
mod.without_c = lm(time ~ rotor * leg_len * clip, data = plane)
anova(mod.without_c, mod.all)
```

```
## Analysis of Variance Table
##
## Model 1: time ~ rotor * leg_len * clip
## Model 2: time ~ rotor * leg_len * leg_wid * clip
##   Res.Df    RSS Df Sum of Sq     F    Pr(>F)
## 1     72 2.3929
## 2     64 1.2547  8    1.1382 7.257 8.313e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Leg Clip

```
mod.without_d = lm(time ~ rotor * leg_len * leg_wid, data = plane)
anova(mod.without_d, mod.all)
```

```
## Analysis of Variance Table
##
## Model 1: time ~ rotor * leg_len * leg_wid
## Model 2: time ~ rotor * leg_len * leg_wid * clip
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     72 2.7479
## 2     64 1.2547  8    1.4932 9.5203 1.446e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the pair-wise results, we can see that the relative importance to flight duration follows: `clip` > `leg_wid` > `rotor` > `leg_len`. However, to answer the question of which main factor contributes the most to longer flight, we need to check the model output below. From the results, only rotor length is contributing a positive estimated effect of 0.02 with the most positive-skewed 95% confidence interval of (-0.15690908, 0.196909085). Therefore, out of the four factors, we think having a high rotor length (8.5 cm) is the most important for making helicopters fly longer.

```
summary(mod.all)
```

```
##
## Call:
## lm(formula = time ~ rotor * leg_len * leg_wid * clip, data = plane)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3800 -0.0665  0.0060  0.0855  0.2320
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     1.92000    0.06262  30.662   <2e-16 ***
## rotor1                          0.02000    0.08855   0.226   0.8220
## leg_len1                       -0.00800    0.08855  -0.090   0.9283
## leg_wid1                       -0.06000    0.08855  -0.678   0.5005
## clip1                          -0.02400    0.08855  -0.271   0.7872
## rotor1:leg_len1                 0.06600    0.12524   0.527   0.6000
## rotor1:leg_wid1                 0.27600    0.12524   2.204   0.0311 *
## leg_len1:leg_wid1               0.00400    0.12524   0.032   0.9746
## rotor1:clip1                    0.20000    0.12524   1.597   0.1152
## leg_len1:clip1                 -0.20200    0.12524  -1.613   0.1117
## leg_wid1:clip1                 -0.25600    0.12524  -2.044   0.0451 *
## rotor1:leg_len1:leg_wid1       -0.29800    0.17711  -1.683   0.0973 .
## rotor1:leg_len1:clip1          -0.04600    0.17711  -0.260   0.7959
## rotor1:leg_wid1:clip1          -0.35200    0.17711  -1.987   0.0512 .
## leg_len1:leg_wid1:clip1         0.08600    0.17711   0.486   0.6289
## rotor1:leg_len1:leg_wid1:clip1  0.29800    0.25047   1.190   0.2385
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.14 on 64 degrees of freedom
## Multiple R-squared:  0.6807, Adjusted R-squared:  0.6058
## F-statistic: 9.094 on 15 and 64 DF,  p-value: 8.164e-11
```

Moreover, if we assume there is no interactions, we fit a linear regression model with main effects only and discovers that high rotor length is still the only treatment that helps with longer flight duration, with an estimated effect of 0.15425 and a 95% confidence interval of (0.08084158, 0.22765842).

```r
# main effect model
mod.main = lm(time ~ rotor + leg_len + leg_wid + clip, data = plane)
summary(mod.main)
```

```
##
## Call:
## lm(formula = time ~ rotor + leg_len + leg_wid + clip, data = plane)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.41887 -0.10275  0.00625  0.11850  0.25837
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.99462    0.04120  48.414  < 2e-16 ***
## rotor1       0.15425    0.03685   4.186 7.65e-05 ***
```

```
## leg_len1    -0.10125    0.03685  -2.748  0.00751 **
## leg_wid1    -0.15175    0.03685  -4.118 9.73e-05 ***
## clip1       -0.19375    0.03685  -5.258 1.33e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1648 on 75 degrees of freedom
## Multiple R-squared:  0.4816, Adjusted R-squared:  0.4539
## F-statistic: 17.42 on 4 and 75 DF,  p-value: 3.803e-10
```

```
confint(mod.main)
```

```
##                   2.5 %      97.5 %
## (Intercept)  1.91255189  2.07669811
## rotor1       0.08084158  0.22765842
## leg_len1    -0.17465842 -0.02784158
## leg_wid1    -0.22515842 -0.07834158
## clip1       -0.26715842 -0.12034158
```

Q2: Interaction Effect

Referring back to the output summary of model **mod.all**, we can see that the interaction term **rotor1:leg_wid1** is significant under $\alpha = 0.05$, with an estimated coefficient of 0.27600 and a 95% confidence interval of (0.02581277, 0.52618723). Therefore, we can say that there is an interaction between rotor length and leg width, or the effect of rotor length differs by leg width.

```
confint(mod.all)["rotor1:leg_wid1",]
```

```
##      2.5 %     97.5 %
## 0.02581277 0.52618723
```

To gain better insights into the characteristics of interaction effect, we brought the analysis to one step further: by fitting three separate models, one with main effects only, one with main effects and two-way interactions, and one with all effects, we conducted pair-wise nested F-tests and evaluate whether some interactions are effective.

```
# model with 2-way interactions
mod.twoway = lm(time ~ rotor + leg_len + leg_wid + clip + rotor:leg_len +
                  rotor:leg_wid + rotor:clip + leg_len:leg_wid + leg_len:clip +
                  leg_wid:clip, data = plane)
```

Main Effects & Two-way Interactions

```
#main effects vs. two-way
anova(mod.main, mod.twoway, test='F')
```

```
## Analysis of Variance Table
##
## Model 1: time ~ rotor + leg_len + leg_wid + clip
```

```
## Model 2: time ~ rotor + leg_len + leg_wid + clip + rotor:leg_len + rotor:leg_wid +
##     rotor:clip + leg_len:leg_wid + leg_len:clip + leg_wid:clip
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     75 2.0368
## 2     69 1.4440  6   0.59283 4.7212 0.0004416 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Two-way Interactions & All Interactions

```
#two-way vs. three-way
anova(mod.twoway, mod.all, test='F')
```

```
## Analysis of Variance Table
##
## Model 1: time ~ rotor + leg_len + leg_wid + clip + rotor:leg_len + rotor:leg_wid +
##     rotor:clip + leg_len:leg_wid + leg_len:clip + leg_wid:clip
## Model 2: time ~ rotor * leg_len * leg_wid * clip
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     69 1.4440
## 2     64 1.2547  5   0.18931 1.9312 0.1014
```

Comparing the F statistics and p-values, we can see that all three- and four-way interactions are insignificant under the testing level of $\alpha = 0.05$. This suggests that the high level interactions are not significantly contributing to the explanatory power, but a larger number of sample size might add more information to the analysis and we cannot be definitive at this point.

Q3: Ideal Combination for Longer Flight

To find the best combination of factors, we derived the average time, in seconds, that a helicopter is expected to fly for all combinations along with each estimate's 95% confidence interval. We found these values by modifying the baseline values so the intercept estimate corresponded with our combination of interest. The 95% confidence interval for the intercept estimate then served as our confidence interval for the expected average flight time of a helicopter with the specified combination of values.

These calculations involved creating four new variables:

- `rotor_low`: signifies low rotor length
- `leg_len_low`: signifies low leg length
- `leg_wid_low`: signifies low width length
- `clip_low`: signifies the absence of a clip

```
plane_all <- plane %>%
  mutate(rotor_low = as.factor(ifelse(rotor == 0, 1, 0)),
         leg_len_low = as.factor(ifelse(leg_len == 0, 1, 0)),
         leg_wid_low = as.factor(ifelse(leg_wid == 0, 1, 0)),
         clip_low = as.factor(ifelse(clip == 0, 1, 0)))
```

Then, a full four-way model was calculated based on the combination of interest. For instance, if we were interested in the combination of low rotor length, low leg length, high leg width, and a leg clip (treatment cd), we would specify the model `time ~ rotor * leg_len * leg_wid_low * clip_low`. This formula gives us a baseline value that corresponds with our combination of interest, and therefore the estimate of the intercept is $E[\bar{y}_{cd}]$.

```
# example for treatment cd
cd = lm(time ~ rotor * leg_len * leg_wid_low * clip_low, data = plane_all)
summary(cd)$coefficients[1,]
```

```
##     Estimate   Std. Error      t value     Pr(>|t|)
## 1.580000e+00 6.261789e-02 2.523241e+01 5.735019e-35
```

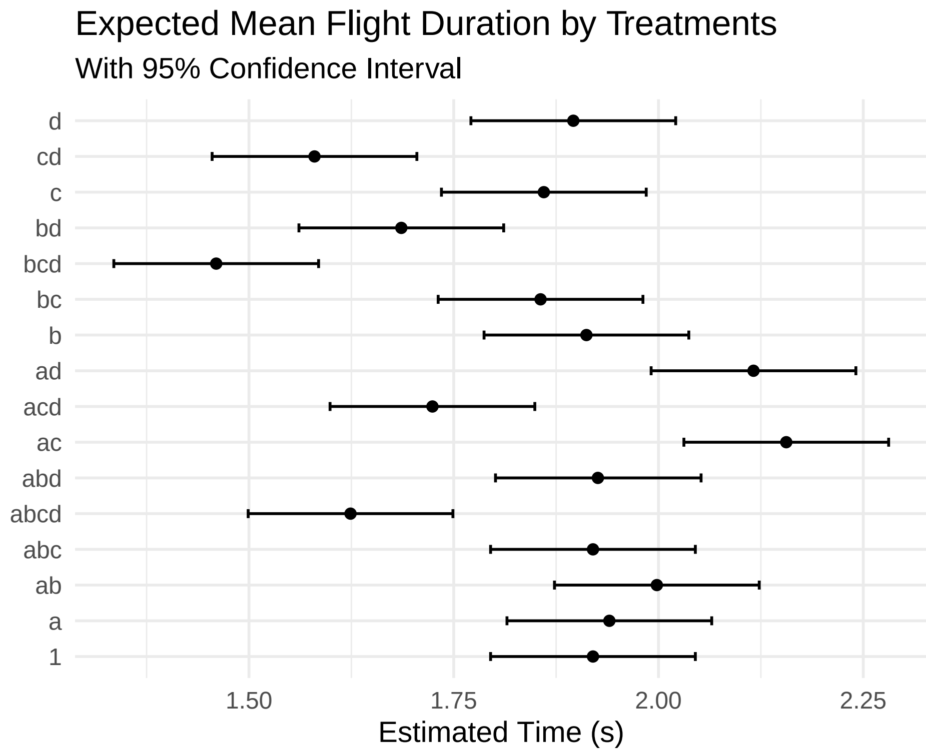
```
# 95% CI for treatment cd (intercept)
confint(cd)[1,]
```

```
##    2.5 %   97.5 %
## 1.454906 1.705094
```

The expected average time in seconds for each helicopter alteration is listed below:

Combination	Lower	Estimate	Upper
1	1.795	1.92	2.045
a	1.815	1.94	2.065
b	1.787	1.912	2.037
c	1.735	1.86	1.985
d	1.771	1.896	2.021
ab	1.873	1.998	2.123
ac	2.031	2.156	2.281
ad	1.991	2.116	2.241
bc	1.731	1.856	1.981
bd	1.561	1.686	1.811
cd	1.455	1.58	1.705
abc	1.795	1.92	2.045
abd	1.801	1.926	2.052
acd	1.599	1.724	1.849
bcd	1.335	1.46	1.585
abcd	1.499	1.624	1.749

```
ggplot(plane_final, aes(x = Estimate, xmin = Lower, xmax = Upper,
                        y = Combination)) +
  geom_errorbarh(height = 0.25) +
  geom_point() +
  theme_minimal() +
  labs(title = "Expected Mean Flight Duration by Treatments",
       subtitle = "With 95% Confidence Interval",
       x = "Estimated Time (s)",
       y = element_blank())
```



As expected from the exploratory data analysis, the helicopters with 8.5cm high rotor length, 7.5cm low leg length, 5cm high leg width, and no leg clip (**treatment ac**) are expected to last longer in the air with an average duration of 2.156 seconds and a 95% confidence interval of (2.031, 2.281). Therefore, we select this combination as the best. However, the combination of high rotor length, low leg length, low leg width, and a clip (treatment ad) was not far behind with an expected average of 2.116 seconds and a 95% confidence interval of (1.991s, 2.241s). Other combinations with overlapping confidence intervals to treatment ac include treatment ab (1.998, (1.873, 2.123)), a (1.94, (1.815, 2.065)), abd (1.926, (1.801, 2.052)), (0) (1.92, (1.795, 2.045)), abc (1.92, (1.795, 2.045)), and b (1.912, (1.787, 2.037)). These all have the potential to perform as well, or even better, than our chosen combinations.